

Method and Apparatus for composition of subtitles

The invention relates to a method and to an apparatus for
5 composition of subtitles for audio/video presentations,
which can be used e.g. for HDTV subtitles in pre-recorded
formats like the so-called Blue-ray Disc.

10 Background

The technique of subtitling for Audio-Visual (AV) material
has been used beginning with the first celluloid cinema
movies and further until the recent digital media appeared.
15 The main target of subtitling has been the support of
handicapped people or small ethnographic language groups.
Therefore subtitling often aims at the presentation of text
information even when having been encoded as graphic data
like pixel maps. Therefore pre-produced AV material for
20 broadcasting (Closed Caption, Teletext, DVB-Subtitle etc.)
and movie discs (DVD Sub-Picture etc.) primarily are opti-
mized for subtitles representing simple static textual in-
formation. However, progress in PC software development for
presentation and animation of textual information induces a
25 corresponding demand for possibilities and features within
the digital subtitling technique used for pre-recording and
broadcasting. Using straightforward approaches without any
special precautions, these increased requirements for subti-
tling would consume a too big portion of the limited overall
30 bandwidth. The conflicting requirements for a 'full feature'
subtitle encompassing karaoke all through genuine animations
are on one hand the coding efficiency and on the other hand
the full control for any subtitle author.

35 For today's state of the art of digitally subtitling AV ma-
terial with separate subtitling information two main ap-
proaches exist: Subtitling can be based on either pixel data

or on character data. In both cases, subtitling schemes comprise a general framework, which for instance deals with the synchronization of subtitling elements along the AV-time axis.

5

Character data based subtitling:

In the character-based subtitling approach, e.g. in the teletext system ETS 300 706 of European analog or digital TV, strings are described by sequences of letter codes, e.g. ASCII or UNICODE, which intrinsically allows for a very efficient encoding. But from character strings alone, subtitling cannot be converted into a graphical representation to be overlaid over video. For this, the intended character set, font and some font parameters, most notably the font size, must either be coded explicitly within the subtitling bitstream or an implicit assumption must be made about them within a suitably defined subtitling context. Also, any subtitling in this approach is confined to what can be expressed with the letters and symbols of the specific font(s) in use. The DVB Subtitling specification ETS 300 743, in its mode of "character objects", constitutes another state-of-the-art example of character-based subtitling.

Pixel data based subtitling:

In the pixel-based subtitling approach, subtitling frames are conveyed directly in the form of graphical representations by describing them as (typically rectangular) regions of pixel values on the AV screen. Whenever anything is meant to be visible in the subtitling plane superimposed onto video, its pixel values must be encoded and provided in the subtitling bitstream, together with appropriate synchronization info, and hence for the full feature animation of subtitles all pixel changed must be transported. Obviously, when removing any limitations inherent with full feature animations of teletext, the pixel-based approach carries the penalty of a considerably increased bandwidth for the subtitling data. Examples of pixel-based subtitling schemes can

be found in DVD's sub-picture concept "DVD Specification for Read-Only disc", Part 3: Video, as well as in the "pixel object" concept of DVB Subtitling, specified in ETS 300 743.

5

Invention

The gist of the invention is a subtitling format encompassing elements of enhanced syntax and semantic to provide improved animation capabilities. The disclosed elements improve subtitle performance without stressing the available subtitle bitrate. This will become essential for authoring content of high-end HDTV subtitles in pre-recorded format, which can be broadcast or pressed on high capacity optical media, e.g. the Blue-ray Disc. The invention includes abilities for improved authoring possibilities for the content production to animate subtitles.

Introduced by the disclosure are elements of syntax and semantic describing the color change for parts of graphics to display. This can be used for highlight effects in applications like for example karaoke, avoiding the repeated transfer of pixel data.

Other disclosed elements of syntax and semantic facilitate the ability of cropping parts of the subtitles before displaying them. By using the technique of subsequently transferred cropping parameters for an object to display, a bit saving animation of subtitles becomes available. Such cropping parameter can be used for example to generate text changes by wiping boxes, blinds, scrolling, wipes, checker boxes, etc.

Furthermore the disclosed elements can be used to provide interactivity on textual and graphical information. Especially the positioning and/or color settings of subtitles can be manipulated based upon user request.

Drawings

Exemplary embodiments of the invention are described with
5 reference to the accompanying drawings and tables, which
show:

- Fig.1: segment_type values for enhanced PCS and RCS;
- Fig.2: Enhanced page composition segment;
- 10 Fig.3: Enhanced region composition segment;
- Fig.4: Example for the definition of a subtitle region and
its location within a page;
- Fig.5: Example for definition of a region sub-CLUT and re-
gion cropping;
- 15 Fig.6: Resulting display example;
- Fig.7: Interactive usage of subtitles;
- Fig.8: Video and Graphics Planes;
- Fig.9: Video and Graphics Mixing and Switching.

20

Exemplary embodiments

The invention can preferably be embodied based on the syntax
and semantic of the DVB subtitle specification (DVB-ST).
25 To provide improved capabilities for the manipulation of
graphic subtitle elements, the semantics of DVB-ST's page
composition segment (PCS) and region composition segment
(RCS) are expanded.

30 DVB-ST uses page composition segments (PCS) to describe the
positions of one or more rectangular regions on the display
screen. The region composition segments (RCS) are used to
define the size of any such rectangular area and identifies
the color-lookup-table (CLUT) used within.

35

The proposed invention keeps backward compatibility with
DVB-ST by using different segment_types for the enhanced PCS

and RCS elements, as listed in Fig.1 showing segment type values according to DVB-ST, with additional values for enhanced PCS and enhanced RCS. It would also be possible to choose other values instead. Another approach for keeping backward compatibility would be to keep the existing segment_types and increase the version_number of the specification, e.g. by incrementing the subtitle_stream_id in the PES_data_field structure.

Fig.2. shows the data structure of an enhanced page composition segment (PCS), containing a region_cropping section and a region_sub_CLUT section. Fig.3 shows the data structure of an enhanced region composition segment (RCS), containing an identifier sub_CLUT_id for a sub-color-look-up-table. With respect to original DVB-ST, all structures shown are expanded. In the tables the additional entries are lines 15-28 in Fig.2 and line 16 in Fig.3.

The enhanced PCS shown in Fig.2 carries optional information about the region cropping and optional information about the region_sub-CLUT for every region listed. The two values of region_cropping and region_sub_CLUT indicate if such optional information is available for the current region in process. Therefore cropping and sub-CLUT may be defined separately for every region. While region_cropping is used as a flag, as indicated by "if region_cropping==0x01", the region_sub_CLUT shows the value how many sub-CLUT positions are described. This is done to provide different alternatives within the stream. Alternative sub-CLUT positions can be used to define different menu button positions for the display screen. Only one of them - the first one as a default - is active and the user can change the position to navigate through the different predefined positions pressing the remote for example.

The enhanced RCS shown in Fig.3 carries the sub_CLUT_id identifying the family of CLUTs that applies to this region. This is done to re-use CLUTs for different regions and dif-

ferent region sub_CLUTs as well.

The enhanced PCS and enhanced RCS elements provide the ability that subtitles can be manipulated independent from the encoding method i.e. independent from whether they are encoded as character data or pixel data.

The enhanced PCS and RCS can be used to perform many different animation effects for subtitles. Those could be wiping boxes, blinds, scrolling, wipes, checker boxes, etc. The following figures show an application example for karaoke. Fig.4 shows the definition of a region R containing lyrics of a song displayed for karaoke. The letters of the subtitle may be encoded as pixel data or as character data as well. The region_vertical_address RVA and the region_horizontal_address RHA define the location of the subtitle within the frame, or page PG, to display.

Fig.5 depicts in the upper part region cropping, and in the lower part the location of the region sub-CLUT. Region cropping defines which part of the region is effectively displayed. This is achieved by four parameters RHC,RVC,RCH,RCW indicating the start coordinates and the size of the fragment to display. region_horizontal_cropping RHC specifies the horizontal address of the top left pixel of this cropping, region_vertical_cropping RVC specifies the vertical address of the top line of this cropping, region_cropping_width RCW specifies the horizontal length of this cropping, and region_cropping_height RCH specifies the vertical length of this cropping, wherein cropping means that part of the subtitles that is visible on a display.

The region sub-CLUT location shown in the lower part of Fig.5 defines which part of the region has to be displayed using a color-look-up-table (CLUT) different from the region CLUT. This is achieved by four parameters SCH,SCVA,SCH,SCW indicating the start coordinates and the size of the sub-region used by the sub-CLUT. All coordinate parameters are

to be understood relative to the region the sub-CLUT belongs to. sub_CLUT_horizontal_address SCHA specifies the horizontal address of the top left pixel of this sub-CLUT, sub_CLUT_vertical_address SCVA specifies the vertical address of the top line of this sub-CLUT, sub_CLUT_width SCW specifies the horizontal length of this sub-CLUT and sub_CLUT_height SCH specifies the vertical length of this sub-CLUT.

Picking up all parameters defined with the previous figures results in the displayed subtitle as depicted in Fig.6. The subtitle is not depicted in whole on the display but only the cropped part of it. Furthermore the sub-CLUT was used to provide a highlight HT, so that the user knows what to sing in the moment.

As the enhanced PCS are sent within MPEG packet elementary stream (PES) packets labeled by presentation time stamps (PTS), any effect can be synchronized to the AV.

20

Another idea of the invention is the superseding of subtitle animation parameters by the user. This offers a way to realize interactive subtitles. The enhanced PCS parameters are transferred as a default, and the user may change them via a remote control for example. Thus the user is able to move, crop or highlight the subtitle.

This could be an advantage for a user defined repositioning of a subtitling text, so that the user can subjectively minimize the annoyance by the subtitle text placement on top of the motion video. Also the color of the subtitles could be set according to users preferences. Fig.7 shows a block diagram for interactive subtitle modifications. The default parameters DD read from a disc D are superseded by superseding data SD being generated upon the user action UA and processed by a processor P.

Another application for overriding subtitle animation parameters like position, cropping rectangle, CLUTs and sub-CLUTs is the realization of some very basic sort of interactive gaming. The subtitle may carry pixel data of an animated character. This character is subsequently moved on the display screen driven by either user interaction, programmatic control or both.

The overriding of subtitle animation parameters can be implemented in at least two ways. The first option is that the overriding parameters SD replace the parameters DD send in the bitstream. The second option is that the overriding parameters SD are used as an offset that is added to or subtracted from the subtitle animation parameters DD send in the bitstream.

The enhanced PCS and RCS provide a lot more of animation capabilities not explained. Following is a non-exhaustive list of examples: wiping boxes, blinds, scrolling, wipes, checker boxes in details.

Exemplary video and graphics planes are shown in Fig.8 in an exemplary, schematic manner. A background is provided by either an MPEG-2 video layer MVL or a still picture layer SPL. They are mutually exclusive, which means that not both of them need to be held in a buffer at a time. The next two layers comprise a subtitle layer SL and an AV sync type graphics layer AVSGL. These two layers are in this example interchangeable, meaning that either the subtitle layer SL or the AV sync type graphics layer AVSGL may have priority over the other. The front layer is a non-AV sync graphics layer NAVSGL, containing graphics that need not be synchronized with the AV content, such as e.g. menus or other on-screen displays. The inventive method can preferably be used for the subtitle layer SL, the AV sync graphics layer AVSGL and/or the Non-AV sync graphics layer NAVSGL.

Fig.9 shows relevant components of an apparatus for video

and graphics mixing and switching. Data comprising either still picture data or MPEG-2 video data, further data for subtitles, data for animations and data for non-AV sync graphics such as menu buttons, are retrieved from a disc D. Additionally or alternatively, data for subtitles, animations and/or non-AV sync graphics can be received from a network NW, e.g. internet. A processing unit CPU processes the non-AV sync graphics data and sends the resulting data to a rendering device for non-AV sync graphics RNAVG.

10 The apparatus contains a still picture decoder SPDec and an MPEG-2 video decoder MVDec, but since only one of them is used at a time, a switch s1 can select which data shall be used for further processing. Moreover, two identical decoders AVSGDec1, AVSGDec2 are used for decoding subtitle and animation data. The outputs of these two decoders AVSGDec1, AVSGDec2 may be switched by independent switches s2, s3 to either a mixer MX, or for preprocessing to a mixer and scaler MXS, which outputs its resulting data to said mixer MX. 15 These two units MX, MXS are used to perform the superimposing of its various input data, thus controlling the display order of the layers. The mixer MX has inputs for a front layer f2, a middle front layer mf, a middle back layer mb and a background layer b2. The front layer f2 may be unused, if 20 the corresponding switch s3 is in a position to connect the second AV sync graphics decoder AVSGDec2 to the mixer and scaler MXS. This unit MXS has inputs for front layer f1, middle layer m and background layer b. It superimposes these data correspondingly and sends the resulting picture data to 25 the background input b2 of the mixer MX. Thus, these data represent e.g. a frame comprising up to three layers of picture and subtitles, which can be scaled and moved together within the final picture. The background input b1 of the mixer and scaler MXS is connected to the switch s1 mentioned 30 above, so that the background can be generated from a still picture or an MPEG-2 video. 35

The output of the first AV sync graphics decoder AVSGDec1 is connected to a second switch s2, which may switch it to the middle layer input m of the mixer and scaler MXS or to the middle back layer input mb of the mixer MX. The output of
5 the second AV sync graphics decoder AVSGDec2 is connected to a third switch s3, which may switch it to the front layer input f1 of the mixer and scaler MXS or to the middle front layer input mf of the mixer MX.

10 Depending on the positions of the second and third switch s2,s3, either the output of the first or the second AV sync graphics decoder AVSGDec1,AVSGD2 may have priority over the other, as described above. For having the data from the first decoder AVSGDec1 in the foreground, the second switch
15 s2 may route the subtitle data to the middle back input mb of the mixer MX, while the third switch s3 routes the animation graphics data to the front input f1 of the mixer and scaler MXS, so that it ends up at the background input b2 of the mixer MX. Otherwise, for having the data from the second
20 decoder AVSGDec2 in the foreground, the switches s2,s3 may route their outputs to the same unit, either the mixer and scaler MXS or the mixer MX, as shown in Fig.9.